

SIS epidemics on Triadic Random Graphs

Ilja Rausch*

Universität Würzburg, Fakultät für Physik und Astronomie, Am Hubland, 97074 Würzburg, Germany

(Dated: September 22, 2016)

It has been shown in the past that many real-world networks exhibit community structures and non-trivial clustering which comes with the occurrence of a notable number of triangular connections. Yet the influence of such connection patterns on the dynamics of disease transmission is not fully understood. In order to study their role in the context of Susceptible-Infected-Susceptible (SIS) epidemics we use the Triadic Random Graph (TRG) model to construct small networks ($N=49$) from distinct, closed, directed triadic subpatterns. We compare various global properties of TRGs and use the N -intertwined mean-field approximation (NIMFA) model to perform numerical simulations of SIS-dynamics on TRGs. The results show that the infection spread on undirected TRGs displays very similar behavior to TRGs with an abundance of (directed) feed-back-loops, while using (directed) feed-forward-loops as network-entities significantly slows down the epidemic and lowers the number of infected individuals in the endemic state. Moreover, we introduce a novel stochastic approach for modelling the SIS-epidemics on TRGs based on characterizing nodes according to their set of (k_{in}, k_{out}) within triads. Within this model, the topology of the network is given by the number and the local structure of directed triadic motifs and not by the adjacency matrix. Nevertheless, the outcome of simulations is qualitatively similar to the results of the NIMFA model.

I. INTRODUCTION

Being connected to other people is an essential part of human life. In the framework of network analysis this phenomenon can be examined from a mathematical point of view. Networks can be found in a vast range of different fields [1–4] and are thus an important part of our society. Furthermore, the interconnectedness of people can be treated scientifically from the perspective of epidemiology. It is obvious that the spread of many airborne and direct contact diseases is influenced by the underlying structure of interpersonal connections [5–7]. The combination of epidemic spreading and network analysis, and deriving corresponding mathematical models is an ongoing scientific challenge [8–11].

One critical question that arises is: how exactly does the structure of a network influence the transmission of a disease? In search for an answer to this question, many different approaches have been suggested, such as consideration of weighted networks [12], dynamic networks [13], generation of networks from real data [14], random and scale-free networks [15–17]. Much attention is devoted to including realistic topological phenomena, for example the existence of community structures (or households) which lead to non-trivial clustering [15, 18, 19]. Such community structures often display a high number of triangular connections, highlighting the importance of accounting for triangles and triadic subgraphs [20, 21]. Moreover, the abundance of loops and specific motifs have been found in numerous real-world networks [22–24]. In this work we examine the role of such substructures on the transmission of diseases. For this purpose

we have constructed ensembles of networks which consist of directed *triadic motifs* [22] as building blocks applying the concept of *Triadic Random Graphs* (TRGs) [25]. Most publications concerned with analysis of epidemic dynamics on complex networks focus on undirected networks [9–11], however numerous real-world networks are either semi-directed or fully directed [11, 26]. Although challenging, including the directionality to our analysis is crucial due to the directional structure of triadic motifs. Moreover, many stochastic approaches to network epidemics consider large networks with a large or infinite number of nodes N [10, 27–29]. However, on a large scale the contribution of local structure of certain motifs may become neglectable [30], whereas on a smaller scale their contribution to the system’s dynamics might be more apparent, which is why small networks, in our case with $N = 49$, might be more appropriate.

In this work we will firstly introduce the concept of Triadic Random Graphs (Sec. II). An extensive set of measurements performed on these graphs will be presented and used for discussions in subsequent sections. We will focus mainly on global properties that are useful in the context of epidemic dynamics. It should be emphasized that whenever TRGs are mentioned, we mean an ensemble of more than 200 samples per type of TRG over which the measured properties and simulations are averaged. However, the samples only differ in the (random) choice of Steiner Triples and not the detailed type of their building blocks, the triadic motifs. In Sec. III we address the question of how the spread of diseases behaves on TRGs with $N = 49$. For this purpose different types of TRGs and their corresponding randomizations were generated. Numerical mean-field simulations of the *Susceptible-Infected-Susceptible* model will be presented. We observe that for most types of triadic subgraphs the fraction of infected individuals x changes with time in a

* ilja.rausch@stud-mail.uni-wuerzburg.de

very similar way. However, TRGs with an abundance of FFL-motifs (\mathcal{T}_1) show a qualitatively different behavior. Furthermore, in order to get a better understanding of epidemic dynamics on TRGs and the influence of triadic motifs, it is worthwhile to consider the stochastic behavior (Sec. IV). After a rough summary of a few existing methods for regular and Erdős-Rényi graphs, we will introduce a novel stochastic model for the epidemic spread on TRGs based on the assignment of characteristic *tags* to the nodes. This node-tagging approach needs much less input than the method used in Sec. III because instead of including the whole topology of a preconstructed network, it focuses only on the type and number of triadic motifs. Remarkably, the numerical results are nonetheless qualitatively similar to those obtained in Sec. III.

II. TRIADIC RANDOM GRAPHS

II.1. Triadic Motifs

One of the most common models for construction of random graphs is the Erdős-Rényi (ER) model [31, 32]. It describes binomial, undirected networks where the number of neighbors, i.e. the total degree k_{tot} , is assigned to every node randomly. Hence, for small networks the probability for a node to have the degree k_{tot} follows the Bernoulli distribution, which converges to Poisson distribution for large networks.

Although ER networks are easy to generate and offer insights into fully random interconnected systems, they may be too simplistic for other scientific purposes and in most cases they fail to include important features of real world networks [33], one of which is clustering, heterogeneous population structures and occurrence of community structures [20, 25, 34, 35]. Furthermore, unlike in ER graphs, in many directed real world networks certain subgraph patterns appear evidently more often than it is expected from randomized equivalents. This was first found by R. Milo et al. [22] who analysed a number of technological and biological data sets and detected certain patterns with above average occurrence frequency which they defined as *network motifs*.

In the following we will focus on a particular kind of directed motifs, the (closed) *triadic motif* \mathcal{M}_i of type i shown in Fig. 1.

II.2. Steiner Triple Systems

An interesting question that arises is: how do triadic motifs influence dynamical processes [36]? In order to answer this question, it is useful to generate and compare networks which differ from each other only through the type of triadic motifs, which play the role of building blocks. So far, there was a notable effort devoted

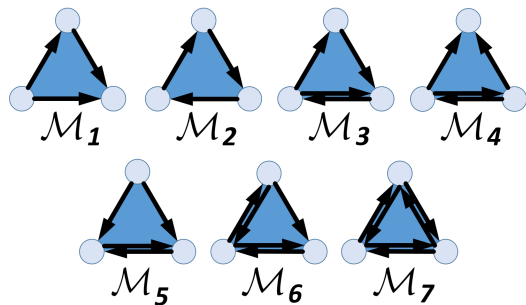


FIG. 1. The seven types of triadic motifs which serve as building blocks for those Triadic Random Graphs \mathcal{T}_i we are concerned with in this article.

to finding a successful method of constructing networks with an abundance of triangles and triads [30, 35, 37, 38]. Recently, one particular model which allows to generate and analyse directed graphs with triadic motifs as primary building blocks has been presented by M. Winkler and J. Reichardt [25]. In their publications the authors made use of the *Steiner Triple System* (STS), which is a mathematical system that consists of distinct subsets of 3 elements (= triples). An important characteristic of the STS is that every pair of elements is connected through exactly one unique link, so that every link can be assigned to exactly one triad. Hence, a graph which is based on Steiner Triples (STs) consists of independent triadic subgraphs. In order to construct a STS with N elements, two necessary and sufficient conditions have to be met [39]:

$$\begin{aligned} N \bmod 2 &= 1 \\ N(N-1) \bmod 3 &= 0 \end{aligned} \quad (1)$$

Moreover, there is an upper bound for the number of dyads-disjoint triads [25]:

$$T \leq \frac{1}{3} = \frac{1}{3} \frac{N(N-1)}{2}, \quad (2)$$

where M is the maximum number of links in a graph.

II.3. Triadic Random Graphs

In analogy to the conditionally independent edges in the random graphs generated by P. Erdős and A. Rényi, the model introduced by M. Winkler and J. Reichardt [25] assumes conditionally independent directed triads and the graphs generated by this model are termed *Triadic Random Graphs* (TRGs). Similar to the ER model, an ensemble of random networks with the same number of nodes and edges can be generated. However, instead of randomly assigning degrees to the nodes, in the TRG model STs are assigned randomly. Hence, by choosing the type and number of triadic motifs displayed in Fig. 1

and assigning them randomly to the nodes, it is possible to construct directed networks with fixed quantities of certain overrepresented triadic motifs and random properties otherwise.

Additionally, to examine their influence on the dynamics of a system, it is necessary to compare TRGs to their randomized counterparts, the *null-models*. The randomization is realized by a Markov Chain Monte Carlo rewiring algorithm [40]. Starting with a TRG sample which is to be randomized, the algorithm performs random but degree-preserving switching of edges between any pair of nodes. Consequently, the triadic structure is lost and ideally there are no overrepresented triadic motifs while the number of nodes, edges and, on average, each node's In- and Out-degrees are preserved. For more detailed information on the graph randomization, see, e.g. [40].

In the present work, TRGs are used to analyse the impact of triadic motifs on a particular kind of dynamics: the epidemic spread of diseases. For this purpose we constructed TRGs purely from conditionally independent triadic motifs of certain kind. A TRG which consists only of triadic motifs of type \mathcal{M}_i will be from now on denoted as \mathcal{T}_i . As mentioned before, every \mathcal{T}_i represents an ensemble of 200 instances of the same kind except for the random distribution of STs. The total degree of every node in such a network has an even value since every node is an element of θ triads ($\theta \in \mathbb{N}$). TRGs are constructed in a way that an arbitrary node ν has the same θ in every \mathcal{T}_i per instance. In other words, for all first instances $\theta_{\mathcal{T}_i, \nu} = \theta_{\mathcal{T}_j, \nu}$. The number of distinct triadic motifs in the resulting TRGs was confirmed using the MFINDER software (version 1.2) [41].

The beauty of TRGs is the fact that their global properties, which are influenced by the abundance of \mathcal{M}_i , often show convenient similarities. Before discussing the results of performed numerical simulations of epidemic spread, it is of importance to look into some of these properties and evaluate their potential impact on the system's dynamics.

II.4. Degree

As TRGs are constructed from STSs, the total degree of each node is even. To every node ν a pair of other nodes is assigned randomly and conditionally independent from other node-pairs, provided that the three nodes are elements of the same ST. Hence, the number of triadic motifs around ν is binomially distributed for small networks and Poisson distributed for large networks. However, the absolute value of the total degree is correlated with the type of the triadic motif. For instance, let us assume that ν is an element of θ triadic motifs and that all motifs are of the type \mathcal{M}_7 (see Fig. 1). Then ν 's total degree is $k_{tot,7} = 2\theta$. However, in the case that all motifs are of the type \mathcal{M}_2 , ν 's total

degree is $k_{tot,2} = \theta = \frac{1}{2}k_{tot,7}$. The total degree of a node ν can be calculated from the adjacency matrix \mathbf{A} : $k_{tot,\nu} = k_{in,\nu} + k_{out,\nu} = \sum_{j=1}^N A_{\nu j} + \sum_{j=1}^N A_{j\nu}$. The adjacency matrix of \mathcal{T}_7 , has exactly two times more entries than the adjacency matrix of \mathcal{T}_1 or \mathcal{T}_2 . Thus, for TRGs which differ only in the choice of \mathcal{M}_i we obtain the relations for the average degrees (averaged over all nodes of the network):

$$\begin{aligned} k_{in,2} &= k_{out,2} = \frac{1}{2}k_{in,7} \\ k_{tot,4} &= k_{tot,1} + k_{bid,4} \\ k_{tot,6} &= k_{tot,1} + 2k_{bid,6} \\ k_{tot,7} &= k_{tot,1} + 3k_{bid,7} = 2k_{tot,1}, \end{aligned} \quad (3)$$

where k_{in} , k_{out} and $k_{bid} = (\mathbf{A}^2)_{\nu\nu}$ denote the in-, out- and bidirected (or reciprocal) degree, respectively. Furthermore, we find the general formulas for the average total degree $\langle k_{tot,i} \rangle$ and the number of edges E_i :

$$\langle k_{tot,i} \rangle = 2 \frac{T}{N} E(\mathcal{M}_i) \quad (4)$$

$$\langle k_{tot,i} \rangle = \frac{E(\mathcal{M}_i)}{E(\mathcal{M}_j)} \langle k_{tot,j} \rangle \quad (5)$$

$$E_i = \frac{E(\mathcal{M}_i)}{E(\mathcal{M}_j)} E_j \quad (6)$$

where $E(\mathcal{M}_i)$ denotes the number of edges within a single triadic motif of type \mathcal{M}_i and E_i the total number of edges within \mathcal{T}_i . Some TRGs have nodes with $k_{in} = 0$ or $k_{out} = 0$, specifically \mathcal{T}_1 , \mathcal{T}_4 and \mathcal{T}_5 , while other \mathcal{T}_i have no such nodes. Denoting $z_{in,i}$ ($z_{out,i}$) the number of nodes with zero in-degree (out-degree) of \mathcal{T}_i , we find that $z_{in,1} = 6.6 \pm 2.1$, $z_{in,5} = 5.4 \pm 2.3$ and zero else, and $z_{out,1} = 5.4 \pm 2.3$ and $z_{out,4} = 5.4 \pm 2.3$ and zero else. The degree distribution depends on the distribution of the triadic motifs in the network and can be calculated from the TRG model. For a more detailed discussion we refer to [25].








II.5. Clustering Coefficient

The clustering coefficient aims to capture quantitatively the fraction of neighbors of a node ν which are also neighbors to each other. Communities are likely to have a high clustering coefficient, whereas it is usually low or zero for random graphs such as the ER graphs. For directed graphs with the adjacency matrix \mathbf{A} the clustering coefficient of node ν is given by [20]:

$$C_\nu(\mathbf{A}) = \frac{\theta_\nu}{T_\nu} = \frac{[(\mathbf{A} + \mathbf{A}^T)^3]_{\nu\nu}}{2k_{tot,\nu}(k_{tot,\nu} - 1) - 2k_{bid,\nu}}, \quad (7)$$

where θ_ν is the number of triangles containing ν and T_ν is the maximum number of triangles that *could* con-

TABLE I. Numerical estimates or various properties of TRGs \mathcal{T}_i which are constructed purely from $T = 60$ closed triadic motifs of type \mathcal{M}_i , averaged over ensembles of >200 samples. Mean values of: total degree $\langle k_{tot} \rangle$, clustering coefficient $\langle C \rangle$, diameter \mathcal{D} , distance $\langle \delta \rangle$, betweenness centrality $\langle B \rangle$, graph spectrum λ assortativity ρ and the current parameter ξ . The brackets $\langle \rangle$ indicate that the values were averaged over all nodes per instance of \mathcal{T}_i . The standard deviation value of k_{tot} is several orders of magnitude smaller than the expectation value and can be therefore neglected. The results in the brackets show $\langle \delta \rangle$ when the infinite values are omitted.

\mathcal{M}_i	$\langle k_{tot} \rangle$	$\langle C \rangle$	\mathcal{D}	$\langle \delta \rangle$	$\langle B \rangle$	λ	ρ	ξ
	7.35	0.176 ± 0.016	∞	$\infty [2.52 \pm 0.80]$	29 ± 14	1.21 ± 0.81	-0.132 ± 0.085	0.885 ± 0.037
	7.35	0.176 ± 0.016	5.73 ± 0.53	2.82 ± 0.35	87.2 ± 1.4	4.41 ± 0.14	0.53 ± 0.14	0.557 ± 0.020
	9.80	0.233 ± 0.029	5.21 ± 0.44	2.56 ± 0.33	75.0 ± 1.4	5.88 ± 0.23	0.43 ± 0.14	0.517 ± 0.018
	9.80	0.233 ± 0.029	∞	$\infty [2.60 \pm 0.52]$	68.1 ± 7.1	5.25 ± 0.29	0.20 ± 0.12	0.515 ± 0.035
	9.80	0.233 ± 0.029	∞	$\infty [2.63 \pm 0.40]$	68.1 ± 7.1	5.25 ± 0.29	0.20 ± 0.12	0.628 ± 0.067
	12.24	0.279 ± 0.031	4.77 ± 0.47	2.36 ± 0.29	65.1 ± 1.2	7.21 ± 0.24	0.38 ± 0.12	0.500 ± 0.017
	14.69	0.352 ± 0.033	4.11 ± 0.31	2.17 ± 0.26	56.3 ± 1.1	8.93 ± 0.29	0.53 ± 0.14	0.500 ± 0.001

tain ν . Note that T_ν includes not only independent triads but rather gives the general number of combinations of how ν 's neighbors can be directly connected to each other. Just as the average degree, the clustering coefficient changes its value depending on \mathcal{M}_i . For the ensembles of \mathcal{T}_i as they are described above (in Sec. II.3), some of the simple relations for the clustering coefficients are:

$$C_{\nu,1} = C_{\nu,2} = \frac{1}{2}C_{\nu,7}, \quad C_{\nu,3} = C_{\nu,4} = C_{\nu,5} \quad (8)$$

The mean values for the clustering coefficient averaged over the whole network $\langle C \rangle$ are shown in Tab. I. It is possible to tune the clustering coefficient within a certain range without changing the total number of triads, e.g. by gradually changing the network from \mathcal{T}_1 to \mathcal{T}_7 . This can be done by replacing one \mathcal{M}_1 by one \mathcal{M}_7 , thus tuning the fraction $\phi(T_1, T_7) = \frac{T_1 - T_7}{T}$, where T_i denotes the number of \mathcal{M}_i triads in the network (unless stated otherwise, for all \mathcal{T}_i we have $T_i = T = 60$).

II.6. Paths and Centrality

Paths and distances on networks are important in the context of disease dynamics. A path between nodes ν_1 and ν_2 represents the sequence of nodes that are passed by while following edges between ν_1 and ν_2 . When individuals are densely connected the disease can spread quickly because of the short paths between the susceptible and infected. There are particularly two kinds of paths that are notable: the shortest path (also known as *distance*) and the longest path (also known as *diameter*). The former considers a path between one pair of nodes

and can be averaged over all pairs to find the mean distance, while the latter gives the maximum longest path between any pair of nodes in the network and is thus a global property. When two nodes are not connected by a path, the length of the path is formally set to ∞ . Hence, the diameter of a network which has more than one connected component is formally infinite, as well as the distance between two nodes of different connected components. At this point, it is worth mentioning that all TRGs used in this article have exactly one connected component. Moreover, on directed graphs the distance between two nodes can still be infinite if there is no path along which all edges are directed. In the numerical example of TRGs with $T = 60$, within \mathcal{T}_1 there are 30.2 ± 4.1 pairs with infinite distances, in \mathcal{T}_4 : 47.2 ± 4.1 and in \mathcal{T}_5 : 6.4 ± 2.8 . The mean distances for different types of TRGs are given in Tab. I. Furthermore, we can provide some relations for the average diameter \mathcal{D}_i of TRGs which differ only in the type of \mathcal{M}_i :

$$\begin{aligned} \mathcal{D}_1 &\rightarrow \infty, \mathcal{D}_4 \rightarrow \infty, \mathcal{D}_5 \rightarrow \infty \\ \mathcal{D}_2 &> \mathcal{D}_3 > \mathcal{D}_6 > \mathcal{D}_7, \end{aligned} \quad (9)$$

where the numbers in the subscripts denote again the type of the overrepresented triadic motifs. Note that for \mathcal{T}_1 , \mathcal{T}_4 and \mathcal{T}_5 the diameter diverges, while for other \mathcal{M}_i the diameter has finite values. The reason for this is associated with reciprocity and directionality of the TRGs. Finally, a relevant measure for the significance of a node ν within a network is the betweenness centrality B_ν [9]. If ν is infected and its B_ν is high, the number of potential infection events during the next iteration step is higher than with a low B_ν .

II.7. Directionality

TRGs of type \mathcal{T}_7 have only bidirectional edges because they are constructed only from subpatterns of type \mathcal{M}_7 . Thus, they can also be considered as undirected. Most publications in the field of complex networks focus on undirected networks [11]. However, the presence of directed edges can influence significantly the dynamics of a system. For instance, the epidemic threshold was found to increase with an increasing fraction of directed to undirected edges on scale-free and binomial networks in SIS NIMFA dynamics [26, 42]. This threshold is equal to $\frac{1}{\lambda}$, where λ is the spectral gap, i.e. the largest real eigenvalue of the adjacency matrix \mathbf{A} . Furthermore, the network's assortativity ρ [43, 44] may have an impact on the spectral gap [26]. The values of λ and ρ for different \mathcal{T}_i are shown in Tab. I.

Moreover, an important global property of TRGs which comes with directionality is the occurrence of a hierarchical structure and, consequently, an inherent directionality. The latter, a term introduced by V. Domínguez-García et al. [23], states that the nodes of a network with a high inherent directionality can be sorted by a hierarchical order so that the network displays an average global direction towards which the nodes preferentially point. This hierarchical ordering can be computed using the equation:

$$h_\nu = 1 + \frac{1}{k_{in,\nu}} \sum_j A_{j\nu} h_j, \quad (10)$$

where h_ν ($k_{in,\nu}$) is the hierarchical level (in-degree), respectively, of the node ν to which node j points. As a boundary condition, h is set to zero for nodes with the lowest value of k_{in} , particularly for those with $k_{in} = 0$. Note that in Eq. 10 it is not always $h_j < h_\nu$, so that in order to examine whether a network shows a hierarchical structure, we calculate the *current parameter* ξ which gives the fraction of edges which are directed from nodes with lower towards nodes with higher h [23] (see Tab. I). Measurements of the current parameter ξ have been performed on over 200 instances of every \mathcal{T}_i with $T = 60$. The highest value of ξ is found for \mathcal{T}_1 . The building blocks of this TRG are also known as *feed-forward-loops* (FFLs). Networks with a high abundance of this kind of motifs are likely to possess a hierarchical structure and high inherent directionality [23]. On the contrast, TRGs with only reciprocal links, \mathcal{T}_7 or undirected triadic graphs, naturally display no preferential direction.

III. EPIDEMICS ON PRECONSTRUCTED TRIADIC RANDOM GRAPHS

III.1. Mean-field approaches

The mathematical foundation of the dynamical process of disease transmission, the *Susceptible-Infected-Recovered* model (SIR), was introduced by W. O. Kermack and A. G. McKendrick [45]. Since then, many optimizations were suggested [11]. Here, we will focus on the individual-based mean-field approach, which offers a simple way to include the network's topology through the adjacency matrix \mathbf{A} [46]. Particularly, we will use the N-intertwined mean field approximation (NIMFA) introduced by P.V. Mieghem [29, 47] for the *Susceptible-Infected-Susceptible* (SIS) model. It is given by a set of ordinary differential equations (ODEs) for the time evolution of the probability x_ν that ν is infected:

$$\frac{dx_\nu}{dt} = \beta(1 - x_\nu) \sum_j A_{j\nu} x_j - \gamma x_\nu. \quad (11)$$

where β is the infection rate and γ the recovery rate. A susceptible individual ν can be infected by other (infectious) individuals j with the probability βdt and subsequently recover (become again susceptible) with the probability γdt . Both processes are assumed to be Poissonian. Eq. 11 can be derived from the Markov theory, using the approximation that the total infection rate of a node is equal to the sum over all probable infection events from its neighbors [48]. Although this model does not include a number of realistic assumptions (e.g. a discrete time-intervals for reinfections [49]), Eq. 11 is computationally simple and it can be straightforwardly used with directed networks, such as Triadic Random Graphs. The fraction of infected individuals x in the population of size N can be calculated using the simple assumption that the total number of infected is $n_I = N \sum_\nu x_\nu$ and hence $x = \frac{n_I}{N} = \sum_\nu x_\nu$.

There are more models for disease transmission that include more realistic assumptions, e.g. that the number of individuals may change due to birth or death, or including further states such as being immune or vaccinated [8]. However, the mathematical complexity increases with more realistic assumptions. In contrast, the SIS-model is rather simple but despite its simplicity it is still being investigated. For instance, an improvement to Eq. 11 is to consider moment closure approximations of second order. This can be done by considering ODEs for pairs on infected nodes [50]. However, this approximation is only valid for sufficiently large networks which is not the case for $N = 49$. Hence, in this section we will mainly focus on the dynamics given by the first-order NIMFA (Eq. 11) and subsequently, in Sec. IV, we will derive a novel stochastic approach which includes closure of higher order.

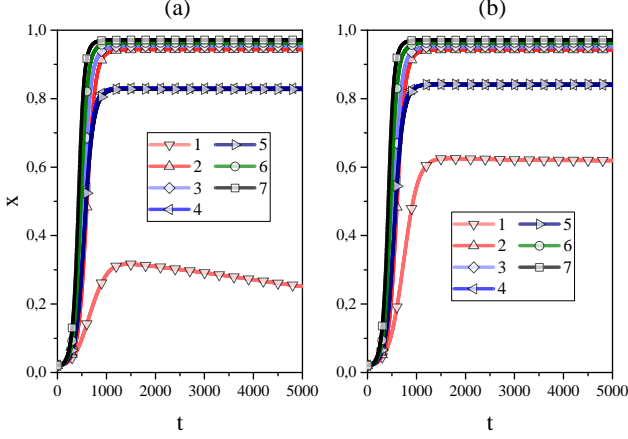


FIG. 2. The change of the fraction of infected individuals x with the number of iteration steps t on (a) different non-randomized TRGs, (b) according null-models. The numbers in the legends correspond to the type of the motif \mathcal{M}_i (see Fig. 1) of which the TRGs consist. Note that the networks differ in nothing but the type of \mathcal{M}_i . The curves of \mathcal{T}_2 , \mathcal{T}_3 , \mathcal{T}_6 , \mathcal{T}_7 coincide nearly perfectly with their null-models. The randomizations of \mathcal{T}_4 and \mathcal{T}_5 have a slightly higher x_{end} than their non-randomized counterparts. The simulation results differ clearly between the original and the randomized versions of \mathcal{T}_1 . Besides the value of x at $t = 5000$, the endemic state sets in for the null-model, on the contrast to the non-randomized \mathcal{T}_1 , where x decreases after reaching a maximum at $t \approx 1300$. Parameters: $N = 49$, $T = 60$, $\beta = 6$, $\gamma = 1$, $x(0) = 0.02$.

III.2. Simulation results

The set of ODEs in Eq. 11 describes how the fraction of infected individuals evolves in time within a population of stationary size. Clustering and triadic substructures can have an impact on their time evolution. Therefore, we performed 4th-order-Runge-Kutta simulations on an ensemble of constructed networks (\mathcal{T}_i) which consist purely of distinctive closed triadic motifs \mathcal{M}_i . For every \mathcal{T}_i we generated over 200 instances, each with $N = 49$ and $T = 60$, over which the simulation results were averaged. Initially, the fraction of infected and susceptible individuals is $x(0) = \frac{1}{49}$ and $s(0) = 1 - x(0)$, respectively. Fig. 2 shows the simulation outcome for the infection rate $\beta = 6$ and the recovery rate $\gamma = 1$.

Except for the curves of \mathcal{T}_1 , \mathcal{T}_4 and \mathcal{T}_5 , all curves reach the endemic state (also known as metastable state, i.e. the non-trivial state where $x = x_{end}$ is constant, before the epidemic starts dying out) between $x_{end} = 0.9$ and $x_{end} = 1.0$. It is remarkable that even though the network is rather small and the triadic motifs clearly differ in their local structure as well as in their influence on the global properties of the networks, as presented in Sec. II, for most \mathcal{T}_i their differences are barely reflected in the simulation outcome. On the other hand, \mathcal{T}_1 , \mathcal{T}_4 and

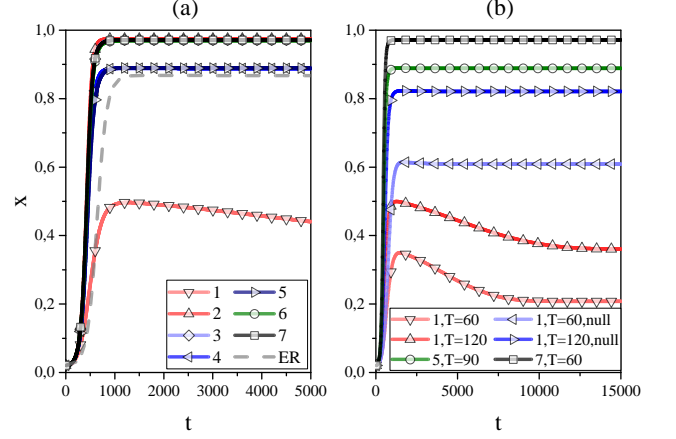


FIG. 3. Mean-field simulation results for the change of the fraction of infected individuals x with the number of iteration steps t on (a) non-randomized \mathcal{T}_i with different total numbers of triads T_i and an Erdős-Rényi graph with $E = 3T_7$, (b) different realizations of \mathcal{T}_1 , \mathcal{T}_5 and \mathcal{T}_7 . All TRGs in (a) have the same total number of edges E . The curves of \mathcal{T}_2 , \mathcal{T}_3 , \mathcal{T}_6 , \mathcal{T}_7 coincide nearly perfectly, same holds for \mathcal{T}_4 and \mathcal{T}_5 . The difference between the upper curves and the curves of \mathcal{T}_4 and \mathcal{T}_5 is approximately the fraction of nodes with $k_{in} = 0$ or $k_{out} = 0$. The simulation result for \mathcal{T}_1 is still clearly separated from the rest and cannot be as simply explained. In (b) results for \mathcal{T}_1 with $T = 60$ and $T = 120$ as well as corresponding null-models are shown on a longer timescale than in (a), curves of \mathcal{T}_5 (\mathcal{T}_7) with $T = 90$ ($T = 60$) are shown for comparison. The endemic state for \mathcal{T}_1 is reached at $t_{end}(T = 60) \approx 9000$ and $t_{end}(T = 120) \approx 13000$. Parameters: $N = 49$, $\beta = 6$, $\gamma = 1$, $x(0) = 0.02$.

\mathcal{T}_5 stand out with $x_{end} \approx 0.85$ and $x(5000) \approx 0.25$, respectively. Furthermore, \mathcal{T}_1 displays a distinctive global maximum before it reaches the endemic state $x_{end} \approx 0.21$ (see Fig. 3). Consequently, a quite noteworthy conclusion is that replacing only one of the edges of the triadic motif \mathcal{M}_1 by a bidirectional edge highly changes the outcome of an epidemic, while this effect is not as severe for other types of \mathcal{M}_i . All these observations hold qualitatively also for networks with lower values of T . For higher T the dynamics approach the behavior of \mathcal{T}_7 .

As one can see in Fig. 2, the diseases spread the fastest (slowest) for \mathcal{T}_7 (\mathcal{T}_1). As discussed in Sec. II.7, \mathcal{T}_7 can be considered as an undirected network. Hence, a susceptible node ν_S can be infected by any of its infectious neighbors and vice versa. The simulation results in Fig. 2 show that, on average, undirected graphs are a good approximation for most TRGs, except \mathcal{T}_1 , \mathcal{T}_4 and \mathcal{T}_5 . As discussed in the previous section, there are clear differences in the number of edges $E = T \cdot E(\mathcal{M}_i)$ between \mathcal{T}_i . By changing the number of triads T we can adjust E according to Eq. 6. Fig. 3 shows that while the curves of \mathcal{T}_2 , \mathcal{T}_3 , \mathcal{T}_6 and \mathcal{T}_7 coincide up to small deviations, they still clearly differ to \mathcal{T}_4 , \mathcal{T}_5 and \mathcal{T}_1 .

Among all TRGs, \mathcal{T}_7 has the lowest diameter and dis-

tance, diseases can spread quickly through the network. Note also that we use only networks with exactly one connected component and the diameter of \mathcal{T}_7 has a finite value, thus all nodes can be reached by an infection. This is not necessarily the case for \mathcal{T}_1 , \mathcal{T}_4 and \mathcal{T}_5 . In fact, for all instances of \mathcal{T}_1 a divergent diameter and several pairs of nodes with a divergent distance are found. Consequently, there are nodes which can never be infected (unless initially), i.e. their in-degree is $k_{in} = 0$ or can never transmit an infection (see also Sec. II.4). Nodes with $k_{in} = 0$ directly reduce x ; nodes with $k_{out} = 0$ can only absorb but not share an infection. Thus, both kinds of nodes do not actively participate in the dynamics of the system and their presence reduces the density of infected individuals x .

Finally, for \mathcal{T}_1 , the far lower value of x_{end} and the occurrence of a global maximum in x is due to the occurrence of hierarchical structure, which we conclude from comparing the simulation outcome to randomized \mathcal{T}_1 in the next section.

III.3. Comparison with Null-Models

Remarkably, considering the relations of the degree distribution or the clustering coefficient between different types of \mathcal{T}_i , we can exclude both properties from contributing significantly to the difference between the simulation outcomes. This becomes most evident when comparing the curves of \mathcal{T}_1 and \mathcal{T}_2 . In order to examine whether other measures might have a greater impact, we compare the results to their corresponding null-models. The lines in Fig. 3 (b) show the simulated curves for randomized TRGs. For every instance of \mathcal{T}_i , 10 randomizations have been generated using the switching algorithm mentioned above (Sec. II.3). Thus, the simulations are averaged over more than 2000 randomizations per \mathcal{T}_i . Except for \mathcal{T}_1 , \mathcal{T}_4 and \mathcal{T}_5 all simulated curves of the null-models coincide perfectly with their non-randomized equivalents. Since the in- and out- degrees of every node are preserved, the number of nodes which cannot be infected or cannot transmit an infection does not change either. Thus, the curves of \mathcal{T}_1 , \mathcal{T}_4 and \mathcal{T}_5 still have a significantly lower x_{end} as compared to other TRGs. However, particularly the dynamics on \mathcal{T}_1 shows clear difference to its randomized counterpart. This remarkable result shows that the local structure of the motifs can in fact have a significant impact on the dynamics and thus influence the epidemic spread without changing N , E and the distribution of the degree or the clustering coefficient. We assume that it is the presence of a hierarchical structure, which is distinctive for networks with an abundance of FFL-motifs [23], that is responsible for the clearly different behavior of epidemic spread on \mathcal{T}_1 . As we will see, this becomes more evident when comparing other measures of \mathcal{T}_1 shown in Tab. I with those of other TRGs,

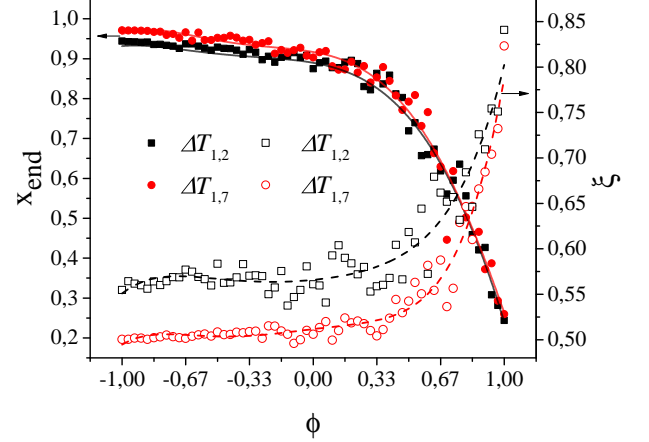


FIG. 4. Bold symbols (**left** y-axis): Change of x_{end} with a decreasing number of \mathcal{M}_2 -motifs (squares) and \mathcal{M}_7 -motifs (circles). Empty symbols (**right** y-axis): Increasing hierarchical structure, represented by the current parameter ξ with a decreasing number of \mathcal{M}_2 -motifs (squares) and \mathcal{M}_7 -motifs (circles). $\Delta T_{i,j} = T_i - T_j$. All lines are guides to the eye.

particularly with \mathcal{T}_2 and \mathcal{T}_7 .

III.4. Fraction ϕ

In order to better understand why x evolves on \mathcal{T}_1 so differently from other \mathcal{T}_i , it is worthwhile to examine how the fraction of infected individuals changes when a TRG approaches the form of \mathcal{T}_1 . Particularly, we're interested in the transition from a purely undirected graph to a graph with an abundance of FFL-motifs, which can be simulated by changing the fraction $\phi(T_1, T_7) = \frac{T_1 - T_7}{60}$, i.e. the numbers of \mathcal{M}_1 and \mathcal{M}_7 triads within the network. We generated TRGs where, starting with 60 \mathcal{M}_7 -motifs (i.e. $T_7 = 60$) out of $T = 60$ triads, i.e. $\phi(0, 60) = -1.0$, we replace one \mathcal{M}_7 -motif by one \mathcal{M}_1 -motif and subsequently perform the same simulations as in Fig. 2 (again, for every value of $\phi(T_i, T_j)$ we generate an ensemble of TRGs with >200 samples). The values of x_{end} in the endemic state for different $\phi(T_1, T_7) = \frac{T_1 - T_7}{60}$ and $\phi(T_1, T_2) = \frac{T_1 - T_2}{60}$ are shown in Fig. 4. The change of x_{end} for a transition from a graph with an abundance of \mathcal{M}_2 (this motif is also known as *feed-back-loop*) to \mathcal{T}_1 is very similar to the transition from \mathcal{T}_7 . In contrast to the latter, \mathcal{T}_2 has only unidirectional links and the total number of edges is $E_2 = E_1$. The fraction of infected individuals in the endemic state decreases for an increasing number of \mathcal{M}_1 in both cases.

The only measures of \mathcal{T}_1 which differ clearly from all other \mathcal{T}_i are: betweenness centrality $\langle B \rangle$, graph spectrum λ , assortativity ρ and current parameter ξ (see Tab. I). However, $\langle B \rangle_1$ has a relatively high deviation from the

expectation value and a high variety of values for every network and even though $\Delta \langle B \rangle_{2,7} = \langle B \rangle_2 - \langle B \rangle_7 > \Delta \langle B \rangle_{2,1} > \Delta \langle B \rangle_{7,1}$ it is not reflected in the values of x_{end} (subscript corresponds to the type of \mathcal{T}_i). Similar arguments suggest that the change of x_{end} cannot be related to λ and ρ . On the contrary, the current parameter ξ_i is close to 0.5 for all \mathcal{T}_i , except \mathcal{T}_5 and \mathcal{T}_1 . As mentioned before (Sec. II.7), when $\xi \approx 0.5$ the network has nearly no hierarchical ordering but when $\xi \approx 1.0$ the network structure is purely hierarchical. Fig. 4 shows the nonlinear increase of ξ for an increasing $\phi(T_1, T_7)$ or $\phi(T_1, T_2)$. Moreover, for the null-model of \mathcal{T}_1 , the value of the current parameter is $\xi_{1,null} = 0.711 \pm 0.068$ and thus lower than for the non-randomized \mathcal{T}_1 . Hence, from the analysis of epidemic dynamics on TRGs we can conclude that within the SIS-model the set in of hierarchical structure, together with the presence of nodes with $k_{in} = 0$ or $k_{out} = 0$, leads to lower fraction of infected individuals.

IV. NODE-TAGGING APPROACH

IV.1. CTMC for regular graphs and Erdős-Rényi-Graphs

In addition to the results of the previous sections, in order to examine whether the local structure of triadic motifs influences the epidemic dynamics, it may be useful to consider an approach which does not include whole preconstructed networks in form of an adjacency matrix. Instead, an investigation from the perspective of a model which focuses only on the type and number of triadic motifs may be advantageous. A realization of such a model is presented in the next subsection (Sec. IV.2). It is based on the Continuous Time Markov Chain (CTMC) and Kolmogorov equations. TRGs which are discussed in this work are relatively small systems, so that neither the number of nodes nor the number of triads can be considered continuous. Furthermore, the spread of diseases throughout all our simulations takes place on static networks, i.e. the topology of our networks does not change with time. This is a good approximation for infections which spread much faster than the change of the network-topology [9]. Thus, when looking at the dynamics of the epidemic we can assume very small time steps. Finally, the epidemic spread is assumed to be a Markov process, i.e. the system has no memory and the dynamics of time $t + dt$ is not influenced by the dynamics at t . With these assumptions (continuous spread of diseases, discrete variables (nodes, triads), Markov assumption) the choice of CTMC is a suitable stochastic model [51]. When the variables can be assumed to be continuous, it is more appropriate to use stochastic differential equations. As a simple example for the analysis of epidemic dynamics we will first discuss the SIS-model on a regular graph, i.e. a graph where all nodes have the same total degree

k_{tot} . A simple deterministic formulation of the change of infected individuals is given by

$$\frac{dn_I}{dt} = \beta n_S n_I - \gamma n_I, \quad (12)$$

where β is the infection rate and γ the recovery rate. Let us denote p_{n_I} the probability of finding n_I infected individuals at time t . The number of the susceptible individuals is then given by $n_S = N - n_I$ (N is the total number of individuals). Within the framework of the CTMC model, p_{n_I} changes with time according to the Kolmogorov forward equations [51]

$$\frac{dp_{n_I}}{dt} = p_{n_I-1} w_{n_I-1}^+ + p_{n_I+1} w_{n_I+1}^- - p_{n_I} (w_{n_I}^+ + w_{n_I}^-), \quad (13)$$

where we used the transition rates $w_{n_I}^+ = \beta n_S n_I$ and $w_{n_I}^- = \gamma n_I$ for the forward and backward transitions, respectively.

The formulation of equations (12) and (13) use the mean-field approximation that the expectation value of the *susceptible-infected pairs* is equal to their product, i.e. $\langle n_{SI} \rangle = \langle n_S \rangle \langle n_I \rangle / N$. This rather ideal assumption can be refined by the *pairwise closure* by deriving differential equations for pairs of individuals, that is for n_{SS} , n_{SI} and n_{II} [27, 52, 53]. Furthermore, stochastic ordinary differential equations for the pairwise-based approach have been presented [54], taking into account that a node's change of state (i.e. becoming infectious or susceptible) quantitatively influences n_{SS} , n_{SI} and/or n_{II} . To elaborate the latter point, let us consider a susceptible node $\nu_S|_{t_1}$ which has k_S susceptible and k_I infected neighbors at time t_1 . Thus, ν_S and its nearest surrounding make up for k_S (SS)-pairs and k_I (SI)-pairs. Consequently, an infection of ν_S (i.e. the transition $\nu_S|_{t_1} \rightarrow \nu_I|_{t_2}$) results in the annihilation of k_S (SS)-pairs ($n_{SS}|_{t_2} = n_{SS}|_{t_1} - k_S$), creation of k_I (II)-pairs ($n_{II}|_{t_2} = n_{II}|_{t_1} - k_I$) as well as annihilation of k_S and creation of k_I (SI)-pairs ($n_{SI}|_{t_2} = n_{SI}|_{t_1} + k_S - k_I$). In other notation we can write

$$\mathbf{v}_{st}|_{t_2} = \mathbf{v}_{st}|_{t_1} + \mathbf{v}_{sh}|_{t_1}, \quad (14)$$

where $\mathbf{v}_{st}|_{t_1} = (n_S, n_I, n_{SS}, n_{SI}, n_{II})$ is the *state vector* at time t_1 and $\mathbf{v}_{sh} = (-1, 1, -k_S, k_S - k_I, k_I)$ the *shift vector* of the forward transition (notation is largely compatible with [28]). For the backward transition \mathbf{v}_{sh} changes its sign. Following this approach, it is important to calculate the probability of finding the central node with exactly k_S susceptible and k_I infected neighbors. In a regular network, where every node has the same total degree k_{tot} and for a susceptible central node ν_S this probability is given by [54]

$$p_{\nu_S}(\mathbf{v}_{st}, \mathbf{v}_{sh}) = \binom{k_{tot}}{k_I} \left(\frac{n_{SI}}{k_{tot} n_S} \right)^{k_I} \left(1 - \frac{n_{SI}}{k_{tot} n_S} \right)^{k_S}. \quad (15)$$

Similarly, the probability for an infected node ν_I is given by

$$p_{\nu_I}(\mathbf{v}_{st}, \mathbf{v}_{sh}) = \binom{k_{tot}}{k_I} \left(\frac{n_{II}}{k_{tot}n_I} \right)^{k_I} \left(1 - \frac{n_{II}}{k_{tot}n_I} \right)^{k_S}. \quad (16)$$

Using these probabilities we can construct the transition rates as

$$w^+(\mathbf{v}_{st}, \mathbf{v}_{sh}) = \beta p_{\nu_S}(\mathbf{v}_{st}, \mathbf{v}_{sh}) \quad (17)$$

$$w^-(\mathbf{v}_{st}, \mathbf{v}_{sh}) = \gamma p_{\nu_I}(\mathbf{v}_{st}, \mathbf{v}_{sh}). \quad (18)$$

The change of pairs with time is then given by

$$\frac{dn_{SI}}{dt} = \sum_{k_I=0}^{k_{tot}} (k_S - k_I) [w^+(\mathbf{v}_{st}, \mathbf{v}_{sh}) - w^-(\mathbf{v}_{st}, \mathbf{v}_{sh})] \quad (19)$$

$$\frac{dn_{II}}{dt} = \sum_{k_I=0}^{k_{tot}} k_I [w^+(\mathbf{v}_{st}, \mathbf{v}_{sh}) - w^-(\mathbf{v}_{st}, \mathbf{v}_{sh})] \quad (20)$$

$$n_{SS} = Nk_{tot} - n_{SI} + n_{II}. \quad (21)$$

This approach can be easily extended from regular to ER graphs by adjusting the probabilities within the transition rates. In ER graphs the distribution of the node's degree is binomial. Hence, the probability of finding the susceptible central node ν_S surrounded by k_S susceptible and k_I infected neighbors can be obtained from the binomial distribution [28]:

$$p_{\nu_S}(\mathbf{v}_{st}, \mathbf{v}_{sh}) = \binom{2n_{SS}}{k_S} \left(\frac{1}{n_S} \right)^{k_S} \left(1 - \frac{1}{n_S} \right)^{2n_{SS}-k_S} \cdot \binom{n_{SI}-1}{k_I-1} \left(\frac{1}{n_S} \right)^{k_I-1} \left(1 - \frac{1}{n_S} \right)^{n_{SI}-k_I}, \quad (22)$$

$p_{\nu_I}(\mathbf{v}_{st}, \mathbf{v}_{sh})$ can be derived accordingly.

Next, we will make use of the fact that \mathcal{T}_i are constructed purely from closed triadic motifs \mathcal{M}_i and derive a novel stochastic approach for this type of graphs.

IV.2. Node-tagging approach for TRGs

In the above subsection, Sec. IV.1, we discussed existing stochastic models which aim to refine the stochastic analysis of epidemic spreading on simple networks by simplifying the consideration of possible effects of local clustering and heterogeneities [28, 53, 54]. There are more approaches in the literature which include stochastic models with pairwise closure as well as higher orders of closure [10, 27, 55]. The approach used in the previous subsection works for large ($N \geq 10^3$) populations and is defined for undirected graphs. However, the networks considered in this article are small ($N = 49$) and directed. In order to consider stochastics of TRGs, we

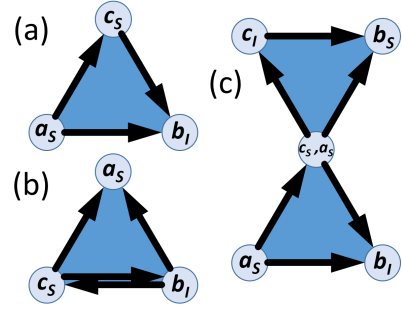


FIG. 5. (a) Feed-Forward-Loop \mathcal{M}_1 . (b) Triadic motif \mathcal{M}_4 . Note that the edge between c_S and b_I is bidirectional, on the contrary to the unidirectional edge in (a). (c) Subgraph constructed from two triadic motifs of type \mathcal{M}_1 with the central susceptible node having a c -tag from the lower motif and an a -tag from the upper motif.

will extend the models discussed in the previous subsection by adjusting the transition rates $w^+(\mathbf{v}_{st}, \mathbf{v}_{sh})$ and $w^-(\mathbf{v}_{st}, \mathbf{v}_{sh})$ accordingly. For this purpose we will introduce a new formulation which is based on assigning characteristic *tags* to nodes of each triadic motif \mathcal{M}_i . Kashtan et al. [56] defined in their work the *role* of a node which describes this node's "uniqueness" within a motif. For triads, two or three nodes have the same role if they can be permuted, while keeping their corresponding (k_{in}, k_{out}) and without changing the triad structure. One can see that there is a clear relation between the role of a node and its in- and out-degrees within a motif. Consider the FFL \mathcal{M}_1 shown in Fig. 5 a). Each node in this triad has a different set of (k_{in}, k_{out}) . The node tagged as a corresponds to the tuple $(k_{in} = 0, k_{out} = 2)$, b to $(2, 0)$ and c to $(1, 1)$. When we add another \mathcal{M}_1 triad, any new nodes are tagged accordingly and the central node gets a second tag (see Fig. 5 (c)). Continuing this approach and extending it to triadic motifs of other types, we introduce 3 rules of the node-tagging (NoTa) approach:

1. Within each triad \mathcal{M}_i there is always exactly one node tagged a , one b and one c .
2. Each tag-index (a, b or c) corresponds to the same set (k_{in}, k_{out}) in every triadic motif of the network.
3. The tags are additive for every node, i.e. for each node of a network there is a corresponding tag-triple $(\kappa_a a, \kappa_b b, \kappa_c c)$ with $\kappa_a, \kappa_b, \kappa_c \in \mathbb{N}_0$.

Using this formalism, we can also assign tags to the nodes of the triad \mathcal{M}_4 , as shown in Fig. 5 (b), even though two nodes have an equal set of (k_{in}, k_{out}) , which does not violate the rules because (k_{in}, k_{out}) need not differ within a triad. Consequently, in the motif \mathcal{M}_7 all three tags correspond to the same set $(k_{in} = 2, k_{out} = 2)$. This approach allows to easily connect the local structure of a motif to the global structure of \mathcal{T}_i . The former can

be described by a motif matrix \mathbf{D}_i (i corresponds to the index of \mathcal{M}_i), where

$$d_{\alpha\beta} = \begin{cases} 1, & \text{if } \alpha \text{ can transmit the infection to } \beta \\ 0, & \text{otherwise} \end{cases},$$

where α and β represent tags. Let's consider for example the triad \mathcal{M}_4 . Without loss of generality we can assign tags to the nodes of this triad as in Fig. 5 (b), in which case $(k_{in}, k_{out})_a \neq (k_{in}, k_{out})_b = (k_{in}, k_{out})_c$. The corresponding motif matrix is

$$\mathbf{D}_4 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}. \quad (23)$$

Using this notation we can define the structure of any other triad. The concept of a motif matrix is not new and has been introduced previously by Milo et al. [56]. Note that $\text{Tr}(\mathbf{D}) = 0$ for each \mathbf{D}_i because we exclude self-loops in all our networks. To complete the depiction of the

triad, we assign a state to its nodes (S for susceptible and I for infected), so that $\mathcal{M}_4 = (a_S b_I c_S)_4$ fully describes the triad shown in Fig. 5 (b). For simplicity, we will call a node in a susceptible state and with a b -tag: a b_S -tag (and similarly for other tags and infected states).

Having defined a notation for the triads we move on to considering the motifs of higher order than triadic motifs. Within the following stochastic model a *high-order-motif* consists of a central node ν which is part of θ triads, i.e. it has 2θ neighbors. As an approximation, we neglect all triads of ν 's neighbors in which ν does not participate. The tags of ν can be now precisely determined by the tags of its neighbors. That is, if ν has exactly one neighbor having the a -tag and exactly one neighbor having the b -tag it immediately means that ν has exactly one tag and it is c . This way it suffices to know the tags of ν 's neighbors which are given by the vector χ in Eq. 24, with $\dim(\chi) = 12$. Furthermore, we are now able to describe our system with a new state vector \mathbf{v}_{st} (see Eq. 25), with $\dim(x) = 8$. From \mathbf{v}_{st} and \mathbf{D} the number of (SI) -pairs, where I points to S can be calculated using Eq. 26.

$$\chi = (\chi_{a_S b_S}, \chi_{a_S c_S}, \chi_{b_S c_S}, \chi_{a_I b_S}, \chi_{a_I c_S}, \chi_{b_I a_S}, \chi_{b_I c_S}, \chi_{c_I a_S}, \chi_{c_I b_S}, \chi_{a_I b_I}, \chi_{a_I c_I}, \chi_{b_I c_I})^T \quad (24)$$

$$\mathbf{v}_{st} = (n_{A_S B_S C_S}, n_{A_S B_S C_I}, n_{A_S B_I C_S}, n_{A_I B_S C_S}, n_{A_S B_I C_I}, n_{A_I B_S C_I}, n_{A_I B_I C_S}, n_{A_I B_I C_I})^T \quad (25)$$

$$n_{SI} = \mathbf{v}_{st} \cdot (0, d_{ca} + d_{cb}, d_{ba} + d_{bc}, d_{ab} + d_{ac}, d_{ca} + d_{ba}, d_{ab} + d_{cb}, d_{ac} + d_{bc}, 0)^T \quad (26)$$

In this notation, $n_{A_S B_I C_S}$ denotes the number of all triads in the network with susceptible a - and c -tags and an infected b -tag. Furthermore, $\chi_{a_S b_I}$ denotes a triad in which ν has a neighbor with an a_S -tag and a neighbor with a b_I tag. As we will see later, the vector χ enables us to include into our stochastic analysis the contribution of both, directionality and local structure, i.e. the type of \mathcal{T}_i .

An infection of the susceptible central node ν leads to the reduction of $n_{A_S B_S C_S}$ by $\chi_{a_S b_S}$, $\chi_{a_S c_S}$ and $\chi_{b_S c_S}$ neighbors of ν . Thus, the first component of the shift vector is $-(\chi_{a_S b_S} + \chi_{a_S c_S} + \chi_{b_S c_S})$. At the same time $\chi_{a_S b_S}$ new $(a_S b_S c_I)$ -triads are created and $(\chi_{c_I a_S} + \chi_{c_I b_S})$ of $(a_S b_S c_I)$ -triads are destroyed. The change of other components of the state vector can be derived in the same way. This procedure results in the change of the state which is given by the shift vector

$$\mathbf{v}_{sh} = \begin{pmatrix} -(\chi_{a_S b_S} + \chi_{a_S c_S} + \chi_{b_S c_S}) \\ \chi_{a_S b_S} - (\chi_{c_I a_S} + \chi_{c_I b_S}) \\ \chi_{a_S c_S} - (\chi_{b_I a_S} + \chi_{b_I c_S}) \\ \chi_{b_S c_S} - (\chi_{a_I b_S} + \chi_{a_I c_S}) \\ \chi_{b_I a_S} + \chi_{c_I a_S} - \chi_{b_I c_I} \\ \chi_{a_I b_S} + \chi_{c_I b_S} - \chi_{a_I c_I} \\ \chi_{a_I c_S} + \chi_{b_I c_S} - \chi_{a_I b_I} \\ \chi_{a_I b_I} + \chi_{a_I c_I} + \chi_{b_I c_I} \end{pmatrix} \quad (27)$$

Additionally, we can track the changes to the total numbers of the infected tags in the network n_{A_I} , n_{B_I} and n_{C_I} because they are given by the components of \mathbf{v}_{st} :

$$\begin{aligned} n_{A_I} &= n_{A_I B_S C_S} + n_{A_I B_I C_S} + n_{A_I B_S C_I} + n_{A_I B_I C_I} \\ n_{B_I} &= n_{A_S B_I C_S} + n_{A_S B_I C_I} + n_{A_I B_I C_S} + n_{A_I B_I C_I} \\ n_{C_I} &= n_{A_S B_S C_I} + n_{A_S B_I C_I} + n_{A_I B_S C_I} + n_{A_I B_I C_I}. \end{aligned} \quad (28)$$

Next, similarly to Sec. IV.1, we define the transition rates $w^+(\mathbf{v}_{st}, \mathbf{v}_{sh})$ and $w^-(\mathbf{v}_{st}, \mathbf{v}_{sh})$ for forward and backward transitions:

$$w^+(\mathbf{v}_{st}, \mathbf{v}_{sh}) = \beta p_{\nu_S}(\mathbf{v}_{st}, \mathbf{v}_{sh}) \quad (29)$$

$$w^-(\mathbf{v}_{st}, \mathbf{v}_{sh}) = \gamma p_{\nu_I}(\mathbf{v}_{st}, \mathbf{v}_{sh}). \quad (30)$$

The probabilities are given by the products of distributions

$$p_{\nu_S}(\mathbf{v}_{st}, \mathbf{v}_{sh}, \chi) = \prod_{\chi_i} p_{\nu_S, \chi_i}(\mathbf{v}_{st}, \mathbf{v}_{sh}, \chi_i) \quad (31)$$

$$p_{\nu_I}(\mathbf{v}_{st}, \mathbf{v}_{sh}, \chi) = \prod_{\chi_i} p_{\nu_I, \chi_i}(\mathbf{v}_{st}, \mathbf{v}_{sh}, \chi_i). \quad (32)$$

Here, $p_{\nu_S, \chi_i}(\mathbf{v}_{st}, \mathbf{v}_{sh})$ gives the probability for ν_S of having χ_i pairs of neighbors of type i . As an example, let us consider the probability that the central node ν_S has $\chi_{a_S b_I}$ neighboring pairs of type $a_S b_I$, i.e. that ν_S has

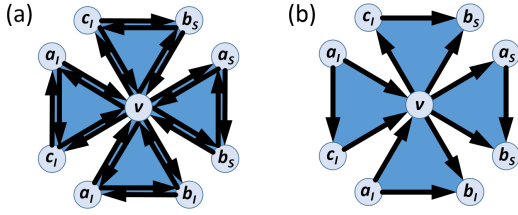


FIG. 6. (a) High-order-motif with 4 \mathcal{M}_7 -triads. (b) High-order-motif with 4 \mathcal{M}_1 -triads. The probability for ν_S to get infected is higher in (a) than in (b).

$\chi_{a_S b_I}$ times the tag c_S , $\chi_{a_S b_I}$ infectious neighbors with the tag b and $\chi_{a_S b_I}$ susceptible neighbors with the tag a . This can be approximated by the probability that out of $n_{A_S B_I C_S}$ -triads we pick the one with the susceptible central node ν_{c_S} exactly $\chi_{a_S b_I}$ -times, which is given by the binomial distribution

$$p_{\nu_S, \chi_{a_S b_I}}(\mathbf{v}_{st}, \mathbf{v}_{sh}, \chi_{a_S b_I}) = \binom{n_{A_S B_I C_S}}{\chi_{a_S b_I}} \left(\frac{1}{C_S}\right)^{\chi_{a_S b_I}} \left(1 - \frac{1}{C_S}\right)^{n_{A_S B_I C_S} - \chi_{a_S b_I}} \quad (33)$$

$$\eta(\mathbf{D}) = \frac{\chi \cdot (0, 0, 0, d_{ac}, d_{ab}, d_{bc}, d_{ba}, d_{cb}, d_{ca}, d_{ac} + d_{bc}, d_{ab} + d_{cb}, d_{ba} + d_{ca})^T}{\sum_{i=4}^9 \chi_i + 2 \sum_{i=10}^{12} \chi_i} \quad (34)$$

Hence, the forward transition rate is now given by

$$w^+(\mathbf{v}_{st}, \mathbf{v}_{sh}) = \beta p_{\nu_S}(\mathbf{v}_{st}, \mathbf{v}_{sh}) \eta(\mathbf{D}) \quad (35)$$

Note that the transition rate, i.e. the recovery process of ν_I , is independent of the type of contact between ν_I and its neighbors. Thus, the rate $w^-(\mathbf{v}_{st}, \mathbf{v}_{sh})$ from Eq. 30 does not change. Within the formalism of the NoTa-approach the fraction $\eta(\mathbf{D})$ makes it strikingly simple to include the type of the TRG-building blocks. Note also, that as in Sec. III we used the assumption of homogeneous rates, i.e. β and γ stay constant throughout the simulation.

IV.3. Numerical results

Fig. 7 shows the evolution of the density of infected population x for different \mathcal{T}_i . Note that the fraction of infected individuals is proportionate to the fraction of infected tags:

$$x = \frac{1}{N} \sum_j^{n_I} \left(\frac{a_{I,j} + b_{I,j} + c_{I,j}}{\kappa} \right), \quad (36)$$

The elements of χ are chosen randomly from uniform distributions. However, as we consider a relatively small system ($N = 49$, $T = 60$), it is important to take into account that at every time step the elements of χ have upper bounds which depend on \mathbf{v}_{st} and the total number of ν 's triads θ . For instance, $\chi_{a_S b_I}$ must not exceed the total number of triads which may include it, i.e. $\chi_{a_S b_I} \leq \min\{(n_{A_S B_I C_S} + n_{A_S B_I C_I}), \theta\}$. Moreover, the rate in Eq. 29 does not entirely cover the contribution of the subgraph structure. Strictly, the infection rate of ν_S depends on the fraction $\eta(\mathbf{D})$ of the infected neighbors which point to ν_S . To elaborate this point, consider the high-order-motif shown in Fig. 6(a), where ν_S has 5 infected neighbors. This network consists of \mathcal{M}_7 -triads, i.e. each infected node can potentially transmit a disease to all of its susceptible neighbors. However, if we replace the triads with motifs of \mathcal{M}_1 -type, as shown in Fig. 6(b), we can see that only 3 of ν_S 's neighbors can transmit a disease to ν_S . Thus, the rate at which ν_S is infected by its neighbors decreases immediately by 3/5, compared to the network with \mathcal{M}_7 -triads. More generally, we can express this impact of the type of \mathcal{M}_i simply by

where κ is the average number of tags per node. Here, we set simply $\kappa = \frac{3T}{N}$, which eventually results in:

$$x = \frac{n_I}{N} = \frac{A_I + B_I + C_I}{3T}. \quad (37)$$

These curves are qualitatively very similar to the curves obtained from mean-field equations shown in Fig. 2. Furthermore, assuming ergodicity we can simulate random network behavior by randomly sampling from different realizations of \mathbf{D}_i . As mentioned in Sec. II the in- and out-degrees of the node do not change for the null-models. This means that ν and its neighbors can still be separated into triples. However, the tags cannot be assigned to any particular set (k_{in}, k_{out}) , so that \mathbf{D}_i can be different at every iteration step. Take the FFL from Fig. 5, there are 6 possible configurations of assigning tags to the nodes, i.e. 6 realizations of \mathbf{D}_i , whereas for a FBL there are only 2. We simulate a randomized network by uniformly sampling from all possible configurations of \mathbf{D}_i . As shown in Fig. 7 (b), there are remarkable quantitative similarities to the mean-field model (Fig. 2 (b)). While the curves of \mathcal{T}_2 , \mathcal{T}_6 , \mathcal{T}_7 show no differences between the randomized and non-randomized realizations, the curves of \mathcal{T}_1 differ clearly.

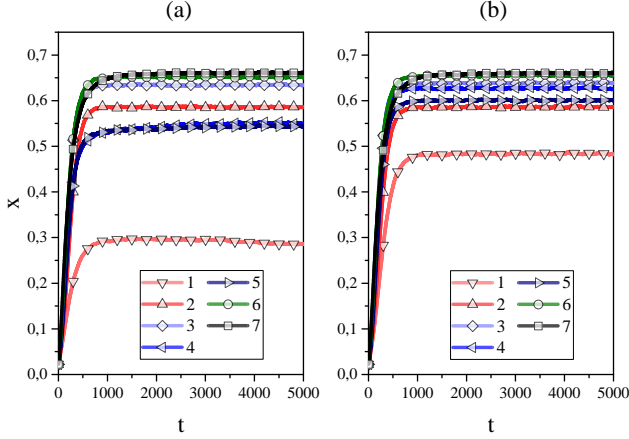


FIG. 7. Numerical solutions to the stochastic model given by Eq. 30 and 35. x is the fraction of infected individuals, t is the number of iteration steps. (a) Non-randomized \mathcal{T}_i , (b) \mathcal{T}_i with randomized motif matrix \mathbf{D} . The results in (a) and (b) are qualitatively similar to the mean-field simulations shown in 2. For \mathcal{T}_1 the number of infected individuals is clearly lower than for other TRGs. In the endemic state, the highest values of x_{end} can be observed on \mathcal{T}_3 , \mathcal{T}_6 and \mathcal{T}_7 and x_{end} for \mathcal{T}_4 and \mathcal{T}_5 lies again between \mathcal{T}_2 and \mathcal{T}_1 . Randomization leads to an increase of the values of x_{end} for \mathcal{T}_4 and \mathcal{T}_5 , as well as a clear increase for \mathcal{T}_1 , while it doesn't change for other TRGs. Parameters: $N = 49$, $T = 60$, $\beta = 6 \cdot 10^3$, $\gamma = 10^3$, $x(0) = 0.02$. The maximum values of θ are: $\theta_1 = \theta_2 = 8$, $\theta_3 = \theta_4 = \theta_5 = 10$, $\theta_6 = 12$ and $\theta_7 = 14$ (number in the subscript corresponds to the index of \mathcal{T}_i).

V. CONCLUSION

Network analysis has become an important tool for the study of various dynamical systems, for instance financial markets, coupled oscillators or epidemics. Examining the properties of real-world networks such as community structures, hierarchical ordering or overrepresentation of specific motifs can lead to new insights on how the system's parameters evolve in time. Supported by the discovery that certain triadic subpatterns occur in real-world data more frequently than expected from random behavior [22], much attention has been devoted to generating networks with an abundance of such motifs and to incorporate triangular subgraphs and triple-closure into the analysis of disease spread. However, numerous growth models assume conditionally independent dyads although this assumption might not be valid for networks with high quantities of triadic patterns [25]. In order to get a better understanding of how the latter might influence the dynamics of the Susceptible-Infected-Susceptible model, we generated ensembles of Triadic Random Graphs \mathcal{T}_i , a special network-class which uses distinct triadic motifs \mathcal{M}_i as building blocks. We compared various global properties of TRGs which differ in the type of their motifs and found many remarkable

equalities as well as differences. For instance, TRGs constructed from feed-forward-loop motifs have the same number of edges, average total degree and average clustering coefficient as TRGs constructed from feed-back-loops but they differ largely in their diameter, mean distance, assortativity and current parameter (i.e. presence of hierarchical structure).

Strikingly, the outcome of simulations on epidemic spread, modelled by the N-intertwined mean-field approximation (NIMFA) [48], shows clear difference between certain types of TRGs. By comparing the simulation results on non-randomized TRGs and the corresponding null-models, we observe that on most \mathcal{T}_i the fraction of infected individuals evolves in time in the same way (up to small deviations). However, the abundance of feed-forward-loops decreases this fraction significantly which is mainly due to the presence of nodes with zero in- or out-degree and the occurrence of a hierarchical structure.

Furthermore, we developed a novel stochastic model based on assigning tags to the nodes of each \mathcal{M}_i according to corresponding sets of (k_{in}, k_{out}) . Instead of including whole preconstructed networks in terms of an adjacency matrix, the only topological parameters used as input are the total number of nodes N and triads T and the type of triadic motifs described by the motif matrix \mathbf{D} . Nevertheless, qualitatively, the simulation outcome resembles the NIMFA curves remarkably well. Hence, the motif structure has clearly an impact on the epidemic dynamics.

These results could improve the understanding of how certain topological patterns change the outcome of an epidemic and how dynamic parameters can be influenced by specific types of networks.

-
- [1] M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001).
 - [2] M. Girvan and M. E. J. Newman, Proceedings of the National Academy of Sciences **99**, 7821 (2002).
 - [3] M. J. Conyon and M. R. Muldoon, Journal of Business Finance & Accounting **33**, 1321 (2006).
 - [4] N. Prulj, D. G. Corneil, and I. Jurisica, Bioinformatics **20**, 3508 (2004).
 - [5] A. S. Klov Dahl, Social Science & Medicine **21**, 1203 (1985).
 - [6] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and W. J. Edmunds, PLoS Med **5**, 1 (2008).
 - [7] V. Colizza, A. Barrat, M. Barthlemy, and A. Vespignani, Proceedings of the National Academy of Sciences of the United States of America **103**, 12015 (2006).
 - [8] K. Rock, S. Brand, J. Moir, and M. J. Keeling, Reports on Progress in Physics **77**, 026602 (2014).
 - [9] L. Danon, A. P. Ford, T. House, C. P. Jewell, M. J. Keeling, G. O. Roberts, J. V. Ross, and M. C. Vernon, "Networks and the epidemiology of infectious disease,"

- (2011).
- [10] J. C. Miller and I. Z. Kiss, *Mathematical Modeling of Natural Phenomena* **9**, 4 (2014).
 - [11] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, *Rev. Mod. Phys.* **87**, 925 (2015).
 - [12] C. Viboud, O. N. Bjørnstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell, *Science* **312**, 447 (2006).
 - [13] M. C. Vernon and M. J. Keeling, *Proceedings of the Royal Society of London B: Biological Sciences* **276**, 469 (2009).
 - [14] J. L. Wylie and A. Jolly, *Sexually transmitted diseases* **28**, 14 (2001).
 - [15] F. Ball, D. Sirl, and P. Trapman, *Mathematical Biosciences* **224**, 53 (2010).
 - [16] A. N. L. M. L. Tom Britton, Maria Deijfen, *Journal of Applied Probability* **45**, 743 (2008).
 - [17] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).
 - [18] M. E. J. Newman and M. Girvan, *Phys. Rev. E* **69**, 026113 (2004).
 - [19] J. V. Ross, T. House, and M. J. Keeling, *PLoS ONE* **5**, 1 (2010).
 - [20] G. Fagiolo, *Phys. Rev. E* **76**, 026107 (2007).
 - [21] S. E. Ahnert and T. M. A. Fink, *Phys. Rev. E* **78**, 036112 (2008).
 - [22] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *Science* **298**, 824 (2002).
 - [23] V. Domínguez-García, S. Pigolotti, and M. A. Muñoz, *Scientific Reports* **4**, 7497 (2014).
 - [24] T. Squartini and D. Garlaschelli, “Triadic motifs and dyadic self-organization in the world trade network,” in *Self-Organizing Systems: 6th IFIP TC 6 International Workshop, IWSOS 2012, Delft, The Netherlands, March 15-16, 2012. Proceedings*, edited by F. A. Kuipers and P. E. Heegaard (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012) pp. 24–35.
 - [25] M. Winkler and J. Reichardt, *Phys. Rev. E* **88**, 022805 (2013).
 - [26] C. Li, H. Wang, and P. Van Mieghem, *Phys. Rev. E* **88**, 062802 (2013).
 - [27] T. House, G. Davies, L. Danon, and M. J. Keeling, *Bulletin of Mathematical Biology* **71**, 1693 (2009).
 - [28] P.-A. Noël, A. Allard, L. Hébert-Dufresne, V. Marceau, and L. J. Dubé, *Journal of Mathematical Biology* **69**, 1627 (2014).
 - [29] P. Van Mieghem, arXiv preprint arXiv:1402.1731 (2014).
 - [30] M. E. J. Newman, *Phys. Rev. Lett.* **103**, 058701 (2009).
 - [31] P. Erdős and A. Rényi, *Publicationes Mathematicae (Debrecen)* **6**, 290 (1959).
 - [32] P. Erdős and A. Rényi, in *Publication of the Mathematical Institute of the Hungarian Academy of Sciences* (1960) pp. 17–61.
 - [33] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
 - [34] M. A. Serrano and M. Boguñá, *Phys. Rev. E* **74**, 056114 (2006).
 - [35] B. Karrer and M. E. J. Newman, *Phys. Rev. E* **83**, 016107 (2011).
 - [36] U. Alon, *Nat Rev Genet* **8**, 450 (2007).
 - [37] P. Holme and B. J. Kim, *Phys. Rev. E* **65**, 026107 (2002).
 - [38] M. Ángeles Serrano and M. Boguñá, *Phys. Rev. E* **72**, 036133 (2005).
 - [39] T. P. Kirkman, *Cambridge and Dublin Math. J* **2**, 1847 (1847).
 - [40] M. Winkler and J. Reichardt, in *2014 IEEE International Conference on Data Mining Workshop* (2014) pp. 605–612.
 - [41] U. Alon, “<https://www.weizmann.ac.il/mcb/urialon/>,” (unpublished).
 - [42] S. C. Ferreira, C. Castellano, and R. Pastor-Satorras, *Phys. Rev. E* **86**, 041125 (2012).
 - [43] M. E. J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
 - [44] M. E. J. Newman, *Phys. Rev. E* **67**, 026126 (2003).
 - [45] W. O. Kermack and A. G. McKendrick, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **115**, 700 (1927).
 - [46] M. J. Keeling and K. T. Eames, *Journal of The Royal Society Interface* **2**, 295 (2005).
 - [47] P. V. Mieghem, *Computer Communications* **35**, 1494 (2012).
 - [48] P. Van Mieghem, *Computing* **93**, 147 (2011).
 - [49] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, *EPL (Europhysics Letters)* **89**, 38009 (2010).
 - [50] E. Cator and P. Van Mieghem, *Phys. Rev. E* **85**, 056111 (2012).
 - [51] M. Keeling and J. Ross, *Journal of The Royal Society Interface* **5**, 171 (2008).
 - [52] M. Newman, *Networks: An Introduction* (Oxford University Press, Inc., New York, NY, USA, 2010).
 - [53] K. T. D. Eames and M. J. Keeling, *Proceedings of the National Academy of Sciences* **99**, 13330 (2002).
 - [54] C. Dangerfield, J. Ross, and M. Keeling, *Journal of The Royal Society Interface* (2008), 10.1098/rsif.2008.0410.
 - [55] E. M. Volz, J. C. Miller, A. Galvani, and L. Ancel Meyers, *PLoS Comput Biol* **7**, e1002042+ (2011).
 - [56] N. Kashtan, S. Itzkovitz, R. Milo, and U. Alon, *Phys. Rev. E* **70**, 031909 (2004).